



**University of
Zurich^{UZH}**

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2017

Adaptive power priors with empirical Bayes for clinical trials

Gravestock, Isaac ; Held, Leonhard

Abstract: Incorporating historical information into the design and analysis of a new clinical trial has been the subject of much discussion as a way to increase the feasibility of trials in situations where patients are difficult to recruit. The best method to include this data is not yet clear, especially in the case when few historical studies are available. This paper looks at the power prior technique afresh in a binomial setting and examines some previously unexamined properties, such as Box P values, bias, and coverage. Additionally, it proposes an empirical Bayes-type approach to estimating the prior weight parameter by marginal likelihood. This estimate has advantages over previously criticised methods in that it varies commensurably with differences in the historical and current data and can choose weights near 1 when the data are similar enough. Fully Bayesian approaches are also considered. An analysis of the operating characteristics shows that the adaptive methods work well and that the various approaches have different strengths and weaknesses.

DOI: <https://doi.org/10.1002/pst.1814>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-146025>

Journal Article

Accepted Version

Originally published at:

Gravestock, Isaac; Held, Leonhard (2017). Adaptive power priors with empirical Bayes for clinical trials. *Pharmaceutical Statistics*, 16(5):349-360.

DOI: <https://doi.org/10.1002/pst.1814>

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Adaptive power priors with empirical Bayes for clinical trials

Isaac Gravestock^{1*}, Leonhard Held¹

Incorporating historical information into the design and analysis of a new clinical trial has been the subject of much discussion as a way to increase the feasibility of trials in situations where patients are difficult to recruit. The best method to include this data is not yet clear, especially in the case when few historical studies are available. This paper looks at the power prior technique afresh in a binomial setting and examines some previously unexamined properties, such as Box p -values, bias and coverage. Additionally it proposes an empirical Bayes type approach to estimating the prior weight parameter by marginal likelihood. This estimate has advantages over previously criticised methods in that it varies commensurably with differences in the historical and current data and can choose weights near 1 when the data are similar enough. Fully Bayesian approaches are also considered. An analysis of the operating characteristics shows that the adaptive methods work well and that the various approaches have different strengths and weaknesses. Copyright © 0000 John Wiley & Sons, Ltd.

Keywords: Bayesian; Clinical trials; Empirical Bayes; Historical data; Power prior

1. Introduction

The incorporation of historical data into clinical trials has become an important topic, as statisticians, sponsors and regulators seek the best ways to continue development in difficult disease areas. Viele et al. [28] have written an excellent overview of the approaches and associated issues. One of the methods considered in that paper is the conditional power prior (with fixed prior weight). This paper is intended to complement that review by looking more closely at the power prior and methods allowing it to be used adaptively which are not discussed by Viele et al.

The power prior is described in a series of articles (Chen, Ibrahim, and Shao [7], Ibrahim and Chen [18], Ibrahim, Chen, and Sinha [20], Chen and Ibrahim [6]; Ibrahim et al. [19]), showing it is a flexible technique and can be applied to a wide range of models, including GLMs and Cox models. Central to the power prior is a weight parameter which determines how much information is used from the historical data. The original formulation is conditional on the weight parameter, but a modification which includes an important normalising factor allows the weight to be treated as unknown [10, 11, 22]. Several authors have criticised the use of unknown weights in power priors [15, 21], claiming they do not adapt appropriately to the difference or similarity between historical and current data. However, we show in this paper that they do behave well given a suitable estimation procedure.

To this end, we propose an empirical Bayes (EB) approach to power prior construction. The main idea of EB is to estimate prior parameters empirically, *i. e.*, from the observed data. It can be considered as hybrid of Bayesian and frequentist thinking and is described clearly in [4]. For power priors, we will use EB methodology to estimate the weight parameter based on the historical and current data. Hobbs et al. [17] present EB estimates for a closely related prior for normal distributions and show their application in generalised linear modelling. Another related approach using EB for adaptive prior weighting is proposed by Held and Sauter [14]. These data-based estimation methods lend themselves to an adaptive trial design, where sample size [13] or randomisation could be adjusted based on the updated prior [16]. For

¹ Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Switzerland

* Correspondence to: E-mail: isaac.gravestock@uzh.ch

comparison, we also examine a fully Bayesian approach to power prior construction. A non-Bayesian alternative is the test-then-pool methodology [28], which compares the historical and current data and, if their difference is not significant at a prespecified level, then the historical and current data are combined in the analysis.

In the context of clinical trials of antibiotics for drug resistant infections, where patients with specific infections can be difficult to recruit, there is often little relevant historical clinical trial information available, which means that we require methods that can appropriately make the best use of whatever information is available. However, we must be careful to acknowledge the evolving resistance and different contexts of historical clinical trials, which means the methodology must handle prior-data conflict appropriately. The FDA draft guidelines for antibacterial therapies for patients with unmet needs [5] discusses the possibility of using results from historical trials and how external data could be incorporated in a down-weighted fashion. We consider power prior methodology to be suitable for this purpose.

1.1. Notation

We construct our informative prior based on the data, x_0 , from a single historical study. The data coming from the current study is denoted x_* . The central assumption of the power prior is that the parameter underlying the historical data, θ_0 , is similar enough to that underlying the current data, θ , that x_0 can be informative about θ . Therefore the priors and posteriors are constructed with likelihoods $L(\theta; x_0)$ and $L(\theta; x_*)$, both around θ . Since this assumption may not be fully met, we control the influence of the historical study on the prior with a weight parameter δ which has a value between 0 and 1.

In clinical trials we are often interested in a binary outcome such as cured/not cured or mortality. In this setting, x_0 and x_* represent the number patients who have an event, while the total number of patients in the historical and current trials are fixed and denoted by n_0 and n_* . The parameter of interest is the success proportion, π , so we use binomial likelihoods, $L(\pi; x_0)$ and $L(\pi; x_*)$. Since we are developing priors based on previous trials, patient-level data may not be available, therefore we only make use of summary-level data. This paper will focus on the cases where the data can be modelled with binomial distributions, but the methodology can be applied more generally. Derivations for normal distributions are given in Appendix A.

2. The Power Prior

If a previous clinical trial has been conducted using the same treatment as that to be used in a new trial, then information from that previous trial may be informative and can be incorporated into the design or analysis of the new trial. The power prior is an especially suitable method of incorporating historical data when only a single previous study available as it does not require the estimation of the heterogeneity between studies [3], unlike a hierarchical model or meta-analysis prior [24, 27].

2.1. Fixed Power Parameter

In its simplest form the power prior is proportional to the likelihood of the historical data raised to the power of the weight $\delta \in [0, 1]$ multiplied by the initial prior $p(\theta)$, *i. e.*

$$p(\theta | x_0, \delta) \propto L(\theta; x_0)^\delta p(\theta). \quad (1)$$

It is therefore a posterior based on the historical data with a modified likelihood. The weighting is a very neat construction because when $\delta = 0$, the likelihood component is 1, and thus the power prior is equal to the initial prior and at the other extreme, when $\delta = 1$, the historical data is used fully, which corresponds to the usual Bayesian updating of the initial prior. This is the original power prior and because it is conditional on the weight parameter δ , it requires that it is known and fixed. The modified power prior which allows the weight to vary is discussed in Section 2.2.

In the binomial setting with unknown proportion π and sample size n_0 , the power prior is

$$p(\pi | x_0, \delta) \propto \binom{n_0}{x_0}^\delta \pi^{\delta x_0} (1 - \pi)^{\delta(n_0 - x_0)} p(\pi).$$

The posterior, based on x_* and n_* , is

$$p(\pi | x_*, x_0, \delta) \propto \binom{n_0}{x_0}^\delta \binom{n_*}{x_*} \pi^{\delta x_0 + x_*} (1 - \pi)^{\delta(n_0 - x_0) + (n_* - x_*)} p(\pi). \quad (2)$$

2.2. Full Bayes

Instead of conditioning on the fixed weight, δ can be treated as an unknown parameter and a prior distribution for it can be included in the formulation. The varying power affects the likelihood such that the constants can no longer be ignored [9]; this requires a normalising constant to be calculated by integrating the original power prior expression over the parameter θ . The joint power prior is then

$$p(\theta, \delta | x_0) = p(\theta | \delta, x_0) p(\delta), \quad (3)$$

where

$$p(\theta | \delta, x_0) = \frac{L(\theta; x_0)^\delta p(\theta)}{\int L(\theta; x_0)^\delta p(\theta) d\theta}. \quad (4)$$

As the weight parameter is limited to $0 \leq \delta \leq 1$, a commonly suggested prior is that $\delta \sim \text{Be}(\alpha, \beta)$. By integrating out the weight parameter, we can find the marginal distribution of θ for fully Bayesian inference [10, 11]. Based on (3), the joint power prior for the event probability π and δ [10] is

$$p(\pi, \delta | x_0) = \frac{\pi^{\delta x_0 + \alpha_\pi - 1} (1 - \pi)^{\delta(n_0 - x_0) + \beta_\pi - 1}}{B(\delta x_0 + \alpha_\pi, \delta(n_0 - x_0) + \beta_\pi)} \text{Be}(\delta | \alpha, \beta), \quad (5)$$

where $B(\cdot, \cdot)$ denotes the beta function and we use a beta prior for the effect parameter, $\pi \sim \text{Be}(\alpha_\pi, \beta_\pi)$. Multiplying (5) by the likelihood of the current data $L(x_\star; \pi)$ gives the joint posterior,

$$p(\pi, \delta | x_0, x_\star) \propto \frac{\pi^{\delta x_0 + x_\star + \alpha_\pi - 1} (1 - \pi)^{\delta(n_0 - x_0) + (n_\star - x_\star) + \beta_\pi - 1}}{B(\delta x_0 + \alpha_\pi, \delta(n_0 - x_0) + \beta_\pi)} \text{Be}(\delta | \alpha, \beta). \quad (6)$$

The marginal posterior distribution of π is

$$p(\pi | x_0, x_\star) \propto \int_0^1 p(\pi, \delta | x_0, x_\star) d\delta.$$

There is no closed form expression due to the beta function in (6), however it can be easily evaluated numerically for full Bayes methodology, which we do in the study of operating characteristics in Section 4. The marginal posterior distribution of δ is found by integrating π out of (6), giving

$$p(\delta | x_0, x_\star) \propto \frac{B(\delta x_0 + x_\star + \alpha_\pi, \delta(n_0 - x_0) + n_\star - x_\star + \beta_\pi)}{B(\delta x_0 + \alpha_\pi, \delta(n_0 - x_0) + \beta_\pi)} \text{Be}(\delta | \alpha, \beta). \quad (7)$$

2.3. Empirical Bayes

An alternative to the fully Bayesian methodology is that of empirical Bayes, which estimates hyper-parameters based on the data. Here we choose the weight parameter with an EB approach, maximising the marginal likelihood of δ ,

$$\hat{\delta}(x_0, x_\star) = \arg \max_{\delta \in [0, 1]} L(\delta; x_0, x_\star), \quad (8)$$

where

$$\begin{aligned} L(\delta; x_0, x_\star) &= \int L(\theta; x_\star) p(\theta | \delta, x_0) d\theta \\ &= \frac{\int L(\theta; x_\star) L(\theta; x_0)^\delta p(\theta) d\theta}{\int L(\theta; x_0)^\delta p(\theta) d\theta}. \end{aligned} \quad (9)$$

Note that $p(\theta | \delta, x_0)$ is properly normalised (as in (4)) and that (9) does not have a prior on δ . Using the empirical Bayes type methodology, we define a new power prior based on the conditional power prior (1) with the estimate (8) of δ inserted, which we will refer to as the EB power prior. The EB estimate eliminates the need for a prior on δ and, by using the marginal likelihood based on x_0 and x_\star , it adapts to the current and historical data. We still require a prior on θ to complete the specification. The EB power prior is still a posterior distribution based on the historical data, but now conditional on $\hat{\delta} = \hat{\delta}(x_0, x_\star)$. In the normal case, the EB estimate of δ is available in closed form, see Appendix A.

The EB estimate of δ for the binomial case comes from maximising the likelihood component in (7), which has the form of a beta-binomial density for x_\star with parameters $\delta x_0 + \alpha_\pi$, $\delta(n_0 - x_0) + \beta_\pi$ and n_\star . There are no closed forms for

the maximum likelihood estimates of the two parameters of a beta-binomial density, but the functions are well behaved enough for numerical estimation. It is necessary to limit the search for the maximum to the interval $[0,1]$ for δ .

Note that the EB estimate $\hat{\delta}$ is not the same as the “guide value” (δ_G) given by Ibrahim et al. [20, 19], which comes from minimizing a penalized likelihood-type criterion,

$$G(\delta) = -2 \log(h(\delta)) + \frac{\log(n_0)}{\delta},$$

$$\text{where } h(\delta) = \int L(\theta; x_*) L(\theta; x_0)^\delta p(\theta) d\theta \quad (10)$$

and n_0 is the sample size of the historical data. This value is based on the original conditional power prior (1) which does not include the normalising constant as in (9), and so suffers from the problems discussed by Neuenschwander et al. [22]. This may be the reason for its counter-intuitive results [20, table 1] where it gives small estimates of δ in the case of perfect agreement in the normal case. Ibrahim et al. [20] stress that this value should only be used as an initial suggestion for the choice of δ and that one should conduct any analysis with several values of δ and compare the results. Figure 1(a) compares the guide values with the empirical Bayes estimates. A further examination of the guide value is given in the normal setting in Appendix A.4. In the binomial setting the guide value does not lead to an estimate of $\delta = 1$, which may be a desirable property if the sample sizes are small and so there is limited information in the likelihood to estimate δ . However this property is not generally present in the guide value, as seen in Figure 5 in Appendix A.4, where in the normal setting the guide value is 1 for a larger range of values than EB.

2.4. Box p -values

Box [2] describes how the prior predictive distribution can be used to test the compatibility of a prior and data. This can be used to identify unsuitable priors that conflict with the observed data. The test gives a p -value that is the probability of X_* being equal or less likely than the observed x_* , in our setting that is

$$\begin{aligned} \text{Box } p(x_* | x_0) &= \Pr\{p(X_* | x_0) \leq p(X_* = x_* | x_0)\} \\ &= \sum_{x=0}^{n_*} p(x | x_0) \mathcal{I}\{p(x | x_0) \leq p(x_* | x_0)\}, \end{aligned}$$

where $\mathcal{I}\{\cdot\}$ is the indicator function. Here $p(x_* | x_0)$ is the prior predictive distribution of the data x_* for the new study, given the historical data x_0 . We note that the modified procedure suggested by Evans and Moshonov [12], is equivalent to that of Box [2] in our case since only summary statistics are used.

The prior predictive distribution for a fixed δ is

$$\begin{aligned} p(x_* | x_0, \delta) &= \int_{-\infty}^{\infty} p(x_* | \theta) L(\theta; x_0, \delta) d\theta \\ &= \text{BeB}(x_* | n_*, \alpha_\pi + \delta x_0, \beta_\pi + \delta(n_0 - x_0)), \end{aligned}$$

where $\text{BeB}(x | n, \alpha, \beta)$ is the beta-binomial density function. For the fully Bayesian approach, it is

$$p(x_* | x_0) \propto \int_0^1 \text{BeB}(x_* | n_*, \alpha_\pi + \delta x_0, \beta_\pi + \delta(n_0 - x_0)) p(\delta) d\delta.$$

Figure 1(b) shows Box p -value curves for historical data $x_0/n_0 = 65/100$ and increasingly differing current data $x_*/200$. Since the power prior is centred around the historical data, all power priors have large p -values when the current data are similar to the historical data. However, as the difference between the current and historical data increases, the Box p -values rapidly drop. The rate of this decrease depends on size of the historical and current studies. In the special case of $\delta = 0$, there is no information in the power prior (given a non-informative initial prior), so there can never be prior-data conflict. The Box p -values for fixed $\delta > 0$ decrease rapidly as x_* increases, meaning that the current data conflicts with the informative power priors.

As the EB prior is controlled by $\hat{\delta}$ based on the current and historical data it never conflicts with the new data. Initially the EB p -values are the same as those from fixed $\delta = 1$ due the upper bound on δ . However, as the difference between x_0/n_0 and x_*/n_* increases, the value of $\hat{\delta}$ decreases, as seen in Figure 1(a), which in turn reduces the information in the prior, thus reducing conflict. As a result of this adjustment the p curve becomes flat. The level of this line is approximately $p = 2(1 - \Phi(1)) \approx 0.32$, which is the same as in the normal setting, see Figure 5 in Appendix A.4. Held and Sauter [14] note similar behaviour of Box p -values when using EB estimates for weights in generalised linear models. The

wiggleness of the lines are due to the discreteness of the beta-binomial distribution. The fully Bayesian approach includes the additional uncertainty in the weight parameter and thus has a larger variance. This means that, with a uniform prior for δ , the posterior predictive distribution adapts weakly to the difference in current and historical data, so there is evidence for prior-data conflict at larger values of x_* . The U-shaped Jeffreys' $\text{Be}(0.5, 0.5)$ prior, which as been suggested for the weight parameter by Dejardin et al. [8], slightly reduces the conflict between the data and the prior by allowing for very little borrowing when the data and prior are dissimilar, thus decreasing further the probability of conflict. This analysis shows the limitations of the fixed approach and how easily it leads to prior-data conflict. It also demonstrates that the fully Bayesian approaches may also exhibit prior-data conflict, depending on the size of the difference and the priors chosen.

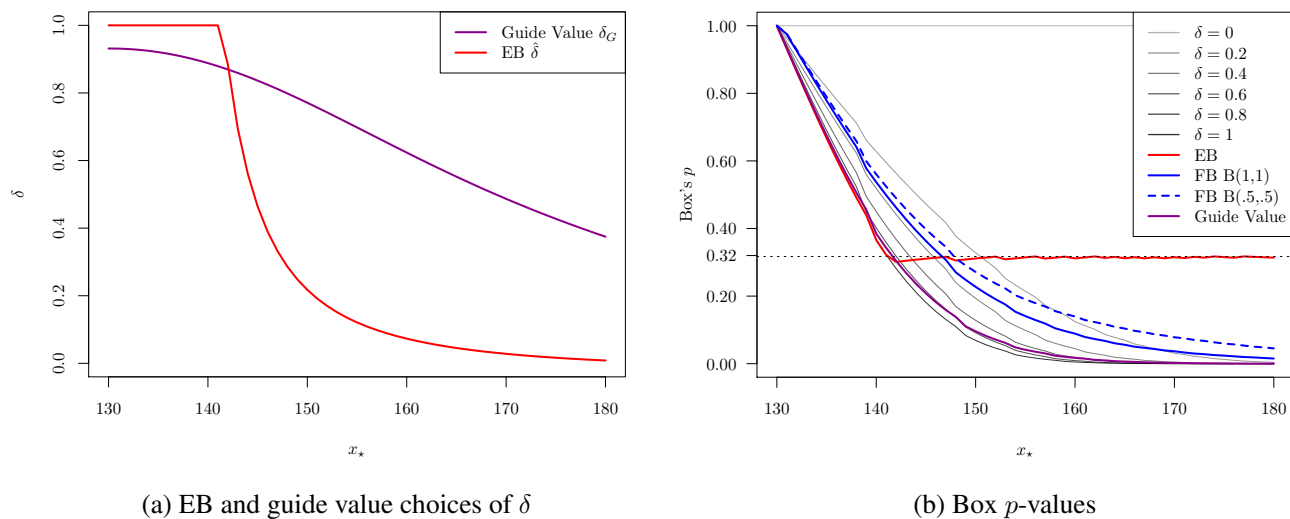


Figure 1. Choices of δ and Box p -values for x_* with $n_* = 200$, $x_0 = 65$ and $n_0 = 100$. The behaviours of the curves in both plots are the same as shown for x_* decreasing from 130.

3. Application

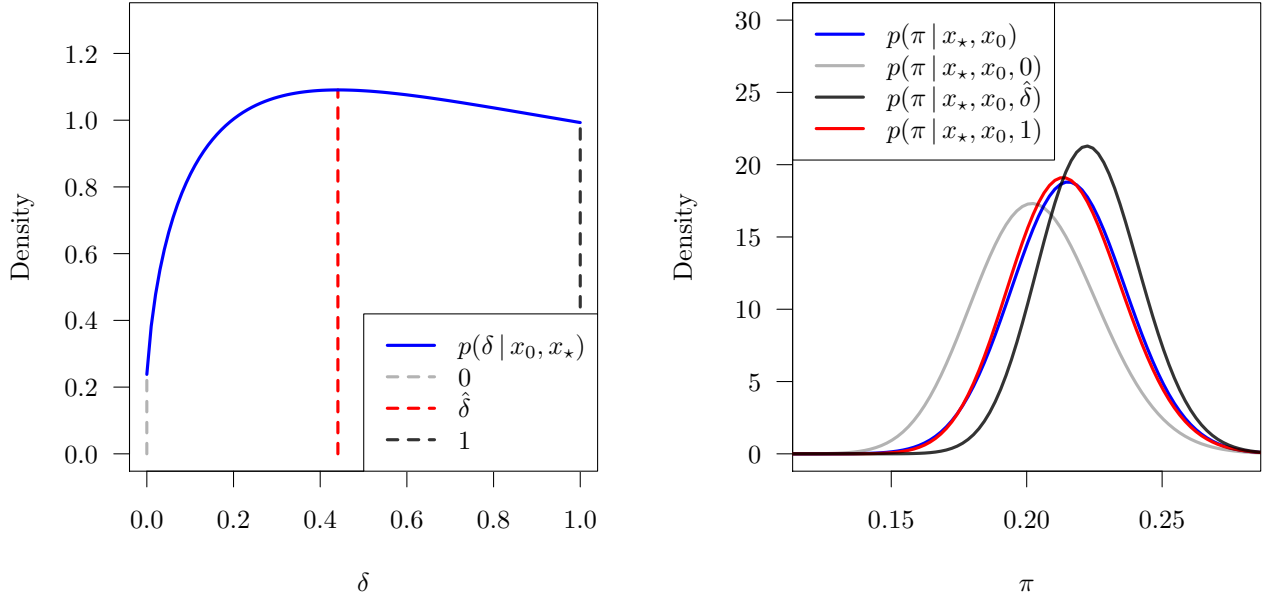
To demonstrate the power prior and estimation by EB, we look at an example based on two clinical trials [25, 29] comparing Linezolid and Vancomycin for the treatment of nosocomial pneumonia. These trials are similar enough that we might expect that the earlier trial could be informative about the latter. Since we are interested in testing the performance of the new treatment, we will only take information about the control treatment, Vancomycin. Table 1 shows the results of the two trials for two outcomes, all cause mortality and cure in the clinically evaluable subset of patients.

Trial	All Cause Mortality	Clinical Cure
Rubinstein (2001)	49/193 (25.4%)	62/91 (68.1%)
Wunderink (2003)	61/302 (20.2%)	111/171 (64.9%)
Risk Difference	5.2% ($p = 0.18$)	3.2% ($p = 0.68$)
95% Confidence Interval	-2.4% to 12.8%	-8.7% to 15.2%

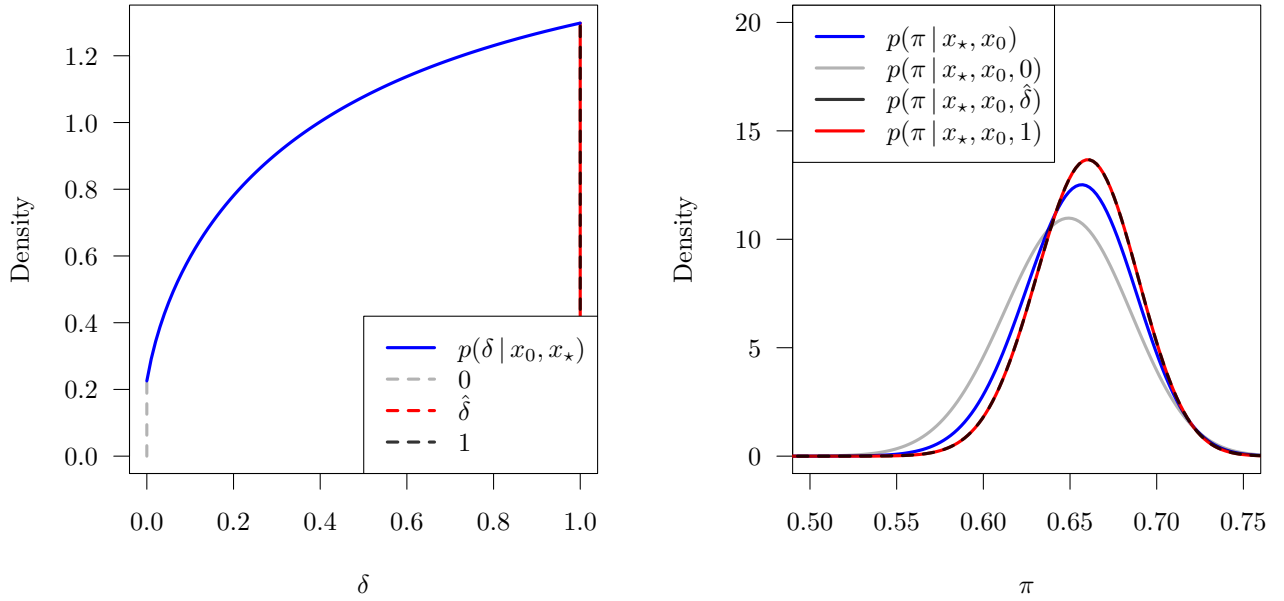
Table 1. Results of Vancomycin arm in trials

First we look at all cause mortality in the two trials. Taking the results of the Rubinstein et al. trial as the historical data and the Wunderink et al. trial as the current data, we can estimate the weight parameter and the posterior distributions of π and δ , as shown in Figure 2(a) where we have used a uniform prior for δ . The EB estimate is $\hat{\delta} = 0.44$ and based on (2), we can interpret the posterior as including $\hat{\delta}n_0$ patients from the historical study in addition to those from the current study. In this case it is equivalent to 85 additional patients. The conditional posterior of $\pi | \hat{\delta}$ under EB is not very different from that of the marginal posterior.

For the cure outcome in the clinically evaluable subset of patients, Figure 2(b) shows the posterior distributions of π and δ . The posterior distribution of δ looks rather different to the previous example. The point estimates of the proportions



(a) Posterior densities based on all cause mortality



(b) Posterior densities based on clinical cure

Figure 2. Marginal posterior densities of π and δ based on historical data from Rubinstein et al. [25] and current data from Wunderink et al. [29].

in each trial are closer together and the sample sizes are also smaller, so there is less evidence of a difference between the trials. This leads to a larger EB estimate, $\hat{\delta} = 1$, meaning 91 extra patients are included in the posterior. This demonstrates that even when the historical and current data are not identical, it is possible to fully use the historical data with the EB

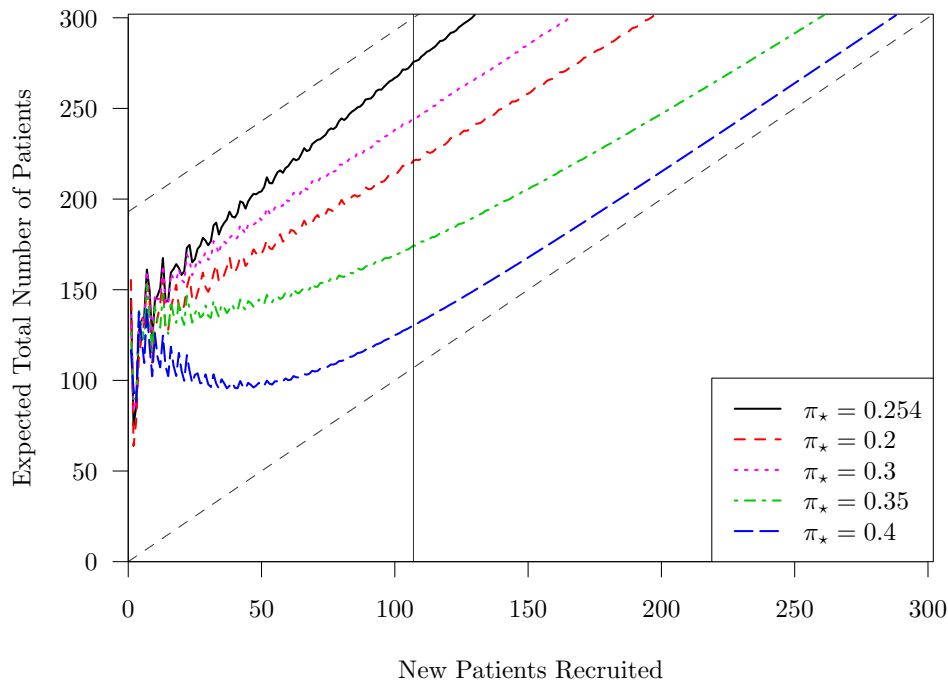


Figure 3. The expected number of patients in the control arm of a trial based on the sum of the current and historical patients for different underlying success probabilities using EB methodology. The upper and lower dashed diagonal lines represent full and no borrowing respectively. An interim look to calculate $\hat{\delta}$ would not be required before the solid vertical line is reached.

approach. There is now more of a difference between the conditional posterior under EB and the marginal posterior.

3.1. Application to Adaptive Trial

Power priors could be used in an adaptive trial design where the historical information is used to supplement the new data. An interim analysis would compare the results in the control arm with the historical study and estimate the weight parameter and thus the number of equivalent patients in the prior. For example, taking the sample sizes from the Wunderink et al. and Rubinstein et al. trials, consider a new trial with a desired sample size of 302, and a prior based on a historical trial with 193 patients. If $\hat{\delta} = 0$ and no information is taken from the historical trial, then 302 new patients would need to be recruited. At the other extreme, if $\hat{\delta} = 1$, then only $302 - 193 = 109$ new patients would be required. Figure 3 shows how the expected total number of patients changes as patients are observed in the new trial for different values of π_* , with the upper and lower bounds shown by dashed lines. The wiggleness is due to the discreteness of the binomial distribution. The expected total number of patients is calculated as

$$\text{Expected Total Patients}(\pi_*, n_*) = n_* + n_0 \sum_{x_*=0}^{n_*} p(x_* | \pi_*) \hat{\delta}(x_0, x_*).$$

It would not be necessary to calculate the weight parameter – and potentially modify any study characteristics – until the minimum possible number of new patients required have been recruited, which in this setting is 109. This allows the value of $\hat{\delta}$ to have stabilised and it can be best seen how many, if any, further patients are required. The EB estimate of the weight parameter could also be used to guide adaptive randomisation in a manner similar to that described by Schmidli et al. [26], and Hobbs et al. [16], where the randomisation ratio is adjusted during the trial. A small δ would mean 1:1 randomisation should continue, but a large weight could allow a greater randomisation ratio to concentrate recruitment into the experimental arm where more evidence is required.

4. Operating Characteristics

In this section we recreate the analysis of Viele et al. [28] and calculate commonly used operating characteristics for evaluating hypothesis tests. The scenario is a clinical trial being conducted with a new treatment (T) and a control treatment (\star) for which a previous clinical trial observed $x_0/n_0 = 65/100$ successful outcomes. In the new trial we have 200 patients in each arm, so we model the outcomes with $X_T \sim \text{Bin}(200, \pi_T)$ and $X_\star \sim \text{Bin}(200, \pi_\star)$. Initial priors are specified for the probability parameters in each arm, $\pi_T, \pi_\star \stackrel{iid}{\sim} \text{Be}(0.001, 0.001)$, for consistency with [28].

The hypotheses to test are $H_0 : \pi_T = \pi_\star$ versus $H_1 : \pi_T > \pi_\star$. The threshold for evidence in favour of H_1 is $\Pr(\pi_T > \pi_\star | x_\star, x_T) > 0.975$, which depends on the two posterior distributions, $\pi_T \sim \text{Be}(0.001 + x_T, 0.001 + 200 - x_T)$ and $\pi_\star \sim \text{Be}(0.001 + 65\delta + x_\star, 0.001 + 35\delta + 200 - x_\star)$.

We compare the mean squared error (MSE), type I error, power, prior sample size, coverage, and bias using power priors based on the EB estimate of δ , fixed values of $\delta = 0, 0.2, 0.4, 0.6, 0.8, 1$ and two fully Bayesian approaches with different initial priors for δ , a uniform $\text{Be}(1, 1)$ and a U-shaped $\text{Be}(0.5, 0.5)$. In addition to the operating characteristics shown in the next section, we compare the power prior approaches with the test-then-pool approach in the Appendix B.

4.1. Expected Prior Sample Size

The prior sample size can be calculated directly from the weight of the historical data. When the initial prior on the parameter is uninformative, the prior sample size is $\delta \times n_0$. The expected number of historical observations incorporated in the priors are shown in Figure 4(a), calculated by taking the expectation over all possible values of X_\star . The EB prior borrows strongly when the new data is similar to the historical data and decreases commensurately as the true value of the parameter moves away. The fixed approaches have constant sample size. The expected prior sample size of the fully Bayesian approaches are calculated as the marginal posterior mean of δ , which shows that they tend to include a moderate number of observations, never including none or all of the historical data. The U-shaped prior shows for a slightly more adaptive behaviour than the uniform prior.

4.2. Mean Squared Error

The MSE is the expected value of the squared error of a point estimate compared with the true value of π_\star , and is calculated as

$$\text{MSE}(\pi_\star | \delta) = \sum_{x_\star=0}^{n_\star} (\pi_\star - \hat{\pi}_\star)^2 \Pr(x_\star | \pi_\star),$$

where $\Pr(x_\star | \pi_\star)$ is the binomial probability mass function. Following Viele et al. [28], we use the posterior mean as this point estimate. For the EB and conditional power priors, we use the posterior mean of the conditional posterior, $p(\pi_\star | x_0, x_\star, \delta)$, which has a beta distribution, so the mean is $\hat{\pi} = (65\delta + x_\star + 0.001)/(100\delta + 200 + 0.002)$. For the FB power prior, we find the mean by numerically integrating over the marginal posterior, $p(\pi_\star | x_0, x_\star)$.

Figure 4(b) shows that all methods that incorporate historical information have the least error when the current study has the same success proportion as the historical data. The MSE of the methods with fixed weighting increase as the true proportion of the new study moves away from the historical proportion. For the adaptive Bayesian methods, the MSE also increases for small differences between historical and true proportion, but as the difference between the historical data and true parameter becomes large their adaptive nature becomes apparent and the MSE begins to decrease.

The $\delta = 0$ curve is relatively flat but shows clearly the decrease in MSE due to the decreasing variance of the binomial distribution with larger probabilities which affects all methods. The $\delta = 1$ curve shows the clear effect on the MSE of the bias increase from inappropriately using historical data at the extremes and of the decreased variance from larger sample size in the centre. The other fixed values behave between these two extreme behaviours. The adaptive methods have similar curves: showing small error near the historical data, increasing as the true parameter differs, before reducing again as the adaptive borrowing reduces the influence of the historical data. Since the EB method adapts most flexibly and borrows the most around the historical data, its MSE increases initially quite sharply. However as the historical data and truth diverge, it corrects sooner than the other methods and decreases again. The full Bayes approach with the uniform prior has larger MSE for $\pi_\star \approx 0.65$ due to borrowing less and increases only slowly for the same reason. It is also slower to adapt to differences between historical and current data leading to larger MSE when the true proportion is far from the historical data. For full Bayes with the U-shaped $\text{Be}(0.5, 0.5)$ prior, we again see that it shows more adaptiveness. It borrows more strongly when there is good compatibility between current and historical data, and reduces borrowing for smaller differences between current and historical data than with the flat prior. However, before the reduction occurs, it borrows very strongly, and so has the largest values for MSE among the adaptive methods considered.

4.3. Type I Error

To examine the type I error of the hypothesis test described earlier, the probability of success in each trial arm is assumed to be equal, $\pi = \pi_* = \pi_T$. The probability of a falsely declaring a difference is

$$\text{Type I Error}(\pi_*) = \sum_{x_*} \sum_{x_T} \Pr(x_* | \pi = \pi_*) \Pr(x_T | \pi = \pi_*) \mathcal{I}\{\Pr(\pi_* < \pi_T | x_*, x_T) > 0.975\}$$

where

$$\Pr(\pi_* < \pi_T | x_*, x_T) = \int_0^1 \int_{\pi_*}^1 p_T(\pi_T | x_T) p(\pi_* | x_*) d\pi_T d\pi_*. \quad (11)$$

Due to the nature of a one-sided test, all methods that incorporate historical information have lower than nominal type I error when the true value is smaller than the historical data and inflated error when the historical data is larger. We are primarily concerned with type I error inflation, which could lead to incorrect conclusions from a clinical trial. The type I errors are shown in Figure 4(c). The adaptive methods have increased error where they borrow, and following the same pattern seen previously, EB has larger error due to its strong borrowing before reducing sharply as it corrects for the difference in historical and current data. The full Bayes approaches are very similar here, with a smaller increase in type I error than EB, but weaker correction as the difference grows large. The $\text{Be}(0.5, 0.5)$ prior again adapts more than the $\text{Be}(1, 1)$, but this only makes an impact when the true parameter far from the historical data.

4.4. Power

Correspondingly, we can calculate the power to detect a difference of 0.12 between the control and experimental treatment with

$$\text{Power}(\pi_*) = \sum_{x_*} \sum_{x_T} \Pr(x_* | \pi_*) \Pr(x_T | \pi_* + 0.12) \mathcal{I}\{\Pr(\pi_* < \pi_T | x_*, x_T) > 0.975\},$$

where $\Pr(\pi_* < \pi_T | x_*, x_T)$ is the same as in (11). The power curves are shown in Figure 4(d). Since we are looking at a one-sided test, the methods can lose power if too much historical information greater than the true parameter is borrowed, but gain power by borrowing when the historical data is less than the true parameter. The $\delta = 0$ curve shows there is a natural increase in power as the true value of π_* increases, due to the decrease in variance of the binomial distribution for more extreme proportions.

In the left third of the plot the intermediate fixed approaches have less power compared to $\delta = 0$ as they are influenced by the historical data which has a larger proportion than the true value, so the difference of 0.12 is harder to detect. The EB approach does not suffer reduced power because the historical data is only used to a small extent. The FB approaches are again similar and have a large decrease in power compared to EB due to their moderate borrowing. In the centre of the plot, where the true parameter and historical data align, all methods that incorporate historical data are more powerful than $\delta = 0$ due to the increased data. EB is most powerful among the adaptive methods because it borrows most strongly.

In the right third of the plot, we again see the effect of the bias induced by borrowing: the power increases as the estimate of π_* is shrunk towards 0.65, therefore making the difference easier to detect. The power of the EB approach flattens as it reduces borrowing. The FB approaches continue increasing in power but tend toward the power level of the $\delta = 0$ curve.

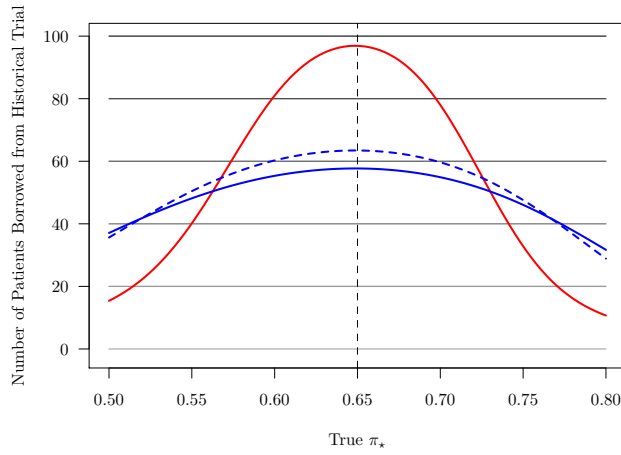
4.5. Coverage

To assess the impact of the priors on the whole distribution of π_* , rather than just the mean, we look at the coverage probabilities. The coverage of 95% equi-tailed credible intervals of the posterior distributions for π_* are shown in Figure 4(e). Specifically, the coverage is computed as

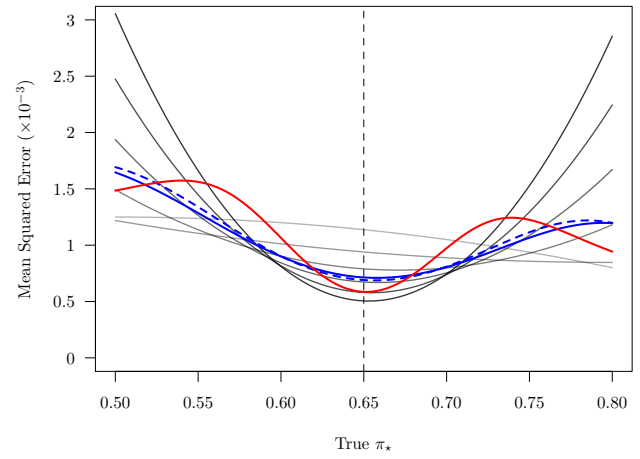
$$\text{Coverage}(\pi_*) = \sum_{x_*} \Pr(x_* | \pi_*) \mathcal{I}\{l(0.025, x_*) < \pi_* < u(0.025, x_*)\},$$

where the upper and lower limits come from the quantile functions, $l(p, x_*) = \{l : \int_{-\infty}^l p(\theta_* | x_*, x_0) = p\}$ and $u(p, x_*) = \{u | u \in (0, 1) : \int_u^{\infty} p(\pi_* | x_*, x_0) = p\}$. For the fixed and EB methods, π_* has a beta distribution with parameters $\delta x_0 + x_* + \alpha_\pi$ and $\delta(n_0 - x_0) + n_* - x_* + \beta_\pi$, so the quantile functions are available. The full Bayes calculations require numerical integration over (6).

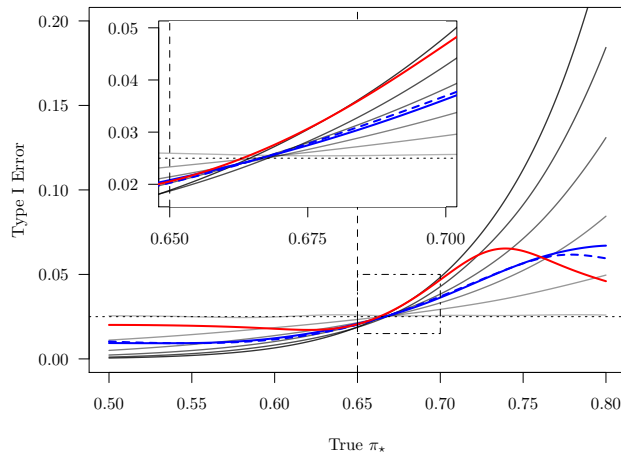
Since the coverage of binomial estimators are unstable, we apply a smoothing procedure [1]. This smoother estimates the local coverage at each point based on beta kernels and we use a smoothing parameter based on the standard deviation of the binomial estimator, $\text{SD}(\hat{\pi}_* = x_*/n_*) = \sqrt{\pi_*(1 - \pi_*)/n_*} \approx 0.03$, assuming π_* is near to the historical data.



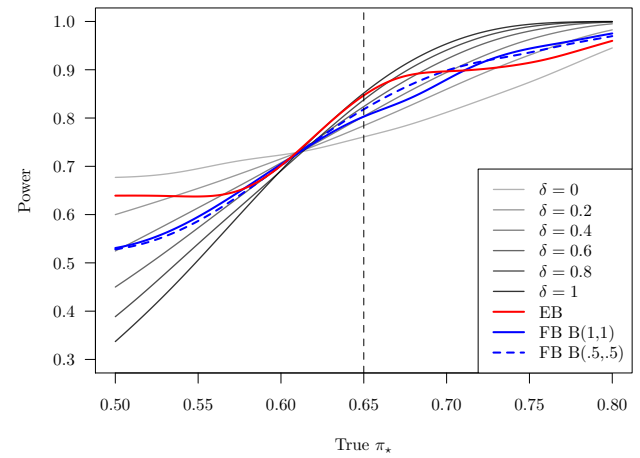
(a) Expected sample size



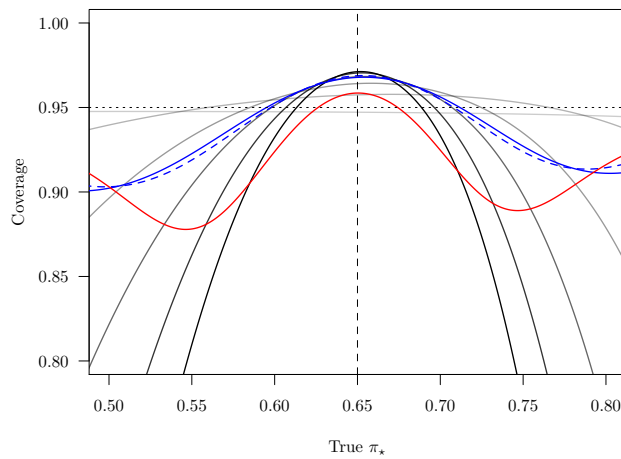
(b) Mean squared error



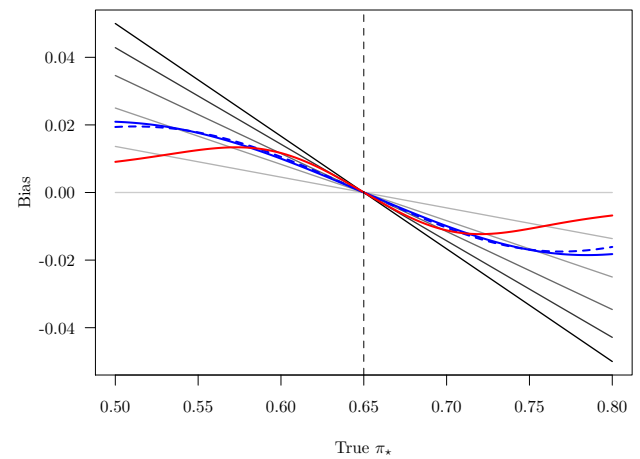
(c) Type I error



(d) Power to detect 0.12 difference



(e) Coverage



(f) Bias

Figure 4. Operating characteristics of fixed, EB and FB methods for historical data $x_0/n_0 = 0.65$ and varying true parameter π_* .

All methods which include the historical data bias the posterior distribution towards the historical data. This means that we observe higher than nominal coverage when the true parameter is near the historical data, which is expected because the credible interval is shifted towards x_0/n_0 and so may include the true parameter, even if the observed value of x_*/n_* is far from π_* and x_0/n_0 . Conversely, when the true π_* is away from the historical data, we see a reduction in the coverage level due to the bias towards the historical data. By using adaptive methods, we expect that there will be less reduction in coverage due to the dynamic reduction in borrowing and therefore bias. The EB and FB priors have similar behaviour, showing a large drop in coverage as the difference between the true π_* and x_0/n_0 increases, before increasing back towards the nominal level for larger differences. The EB prior drops sharply to less than 90% before increasing back towards the nominal level. The FB approaches decrease less sharply, with their minimum coverage always larger than 90%. They increase again, more slowly than EB, with the $\text{Be}(0.5, 0.5)$ prior increasing slightly faster.

4.6. Bias

Complementary to the examination of the MSE, it is useful to look at the bias in the parameter estimates. Here bias is calculated as the expected difference of the posterior mean from the true value, $\mathbb{E}_{X_* | \pi_*}(\pi_* - \tilde{\pi}_*)$. As for the MSE, the mean comes from the conditional posterior for EB and conditional power priors and from the marginal posterior for FB. Figure 4(e) shows that there is negligible difference between the level of bias between the adaptive methods when the true parameter is near the historical data. However as the difference grows the EB method's more sensitive adaptiveness reduces the weight of the historical data to reduce the bias.

5. Discussion

The power prior is a flexible method for specifying an informative prior: it calculates a posterior distribution from a prior and the weighted likelihood of some previous data. This posterior is then used as the prior for a new study. The major criticism of the power prior has been the difficulty in choosing the weight parameter. We have proposed an empirical Bayes approach that estimates the weight parameter based on the maximum of the marginal likelihood given the new and old data. We compared this to the fully Bayesian approach of integrating out the weight parameter, as well as to fixed weights.

The EB approach takes care that the power prior is always compatible with the current data, as seen by the Box p -values never becoming small. It works because the EB estimate reduces the level of information in the prior when necessary. The fully Bayesian approach is adaptive by nature, but does not react as strongly to the differences between the current and historical data. Therefore we see that with fixed weight $\delta > 0$ or the fully Bayesian approach, the power prior methodology can lead to priors that conflict with the current data.

A considerable benefit of the EB estimation is that it can give a weight of 1 to the historical study, which is not the case for the fully Bayesian approach, which, even for identical historical and current data, has posterior mean $\delta \approx 0.6$ for $\delta \sim \text{Be}(1, 1)$ *a priori*. The fully Bayesian approach can be induced to give larger weights by using strong priors on δ , but this would require specific justification. A general criticism of EB is that it does not account for any uncertainty in the parameters it estimates, in this case the weight parameter δ . The greatest impact of this is when the data are similar and EB estimates a large weight, leading to a narrower posterior than with the fully Bayesian approach. This trade off is necessary for the good adaptive performance, and the methodology still accounts for the uncertainty in the parameter of interest θ_* .

A necessary condition for the implementation of the adaptive trial described in Section 3.1 is the comparison of the outcomes of the current patients with those of the historical patients. This requires that the current control arm data be unblinded to whomever is responsible for determining the weight parameter. Many adaptive designs also require such unblinding, so whether this is problematic depends on the context surrounding the design of the trial. The EB power prior methodology provides a reasonable approach to incorporating down-weighted historical data into clinical trials, and if unblinding is not possible the FB approach could be applied.

In case there is more than one relevant historical study available, then it may be appropriate to include them separately in the power prior and have different weights for each. The literature on choosing multiple power prior parameters is fairly sparse; with many papers making a brief remark as to its possibility and difficulty. Notably, Duan [9, ch. 4] presents an example using a fully Bayesian approach. We have chosen not to include another such section and look to prepare a more complete treatment in a future paper. Power priors with individually weighted studies present challenges both mathematically and computationally. The EB approach is, however, feasible. We believe that power priors are still an attractive method despite the challenges, as the weight parameter is very interpretable for compatibility of studies and sample size when compared to the between study heterogeneity parameter in a hierarchical model.

Acknowledgments

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement n^o 115523 [Combatting Bacterial Resistance in Europe – COMBACTE], resources of which are composed of financial contribution from the European Union's 7th Framework Programme (FP7/2007-2013) and European Federation of Pharmaceutical Industries and Associations (EFPIA) companies' in kind contribution.

References

- [1] Bayarri, M. J. and Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science* **19**, 58–80.
- [2] Box, G. E. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society. Series A (General)* **143**, 383–430.
- [3] Campbell, G. (2011). Bayesian statistics in medical devices: Innovation sparked by the FDA. *Journal of Biopharmaceutical Statistics* **21**, 871–887.
- [4] Carlin, B. P. and Louis, T. A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.
- [5] Center for Drug Evaluation and Research (FDA) (2013). Guidance for Industry: Antibacterial Therapies for Patients With Unmet Medical Need for the Treatment of Serious Bacterial Diseases.
- [6] Chen, M.-H. and Ibrahim, J. G. (2006). The relationship between the power prior and hierarchical models. *Bayesian Analysis* **1**, 551–574.
- [7] Chen, M.-H., Ibrahim, J. G., and Shao, Q.-M. (2000). Power prior distributions for generalized linear models. *Journal of Statistical Planning and Inference* **84**, 121–137.
- [8] Dejardin, D., van Rosmalen, J., and Lesaffre, E. (2014). Including historical data in the analysis of clinical trials using the modified power prior. Presentation at Bayes Pharma meeting 2014, <http://www.bayes-pharma.org/bayes2014docs/Day1/Dejardin.pdf>.
- [9] Duan, Y. (2005). *A Modified Bayesian Power Prior Approach with Applications in Water Quality Evaluation*. PhD thesis, Virginia Polytechnic Institute and State University.
- [10] Duan, Y., Smith, E. P., and Ye, K. (2006). Using power priors to improve the binomial test of water quality. *Journal of Agricultural, Biological, and Environmental Statistics* **11**, 151–168.
- [11] Duan, Y., Ye, K., and Smith, E. P. (2006). Evaluating water quality using power priors to incorporate historical information. *Environmetrics* **17**, 95–106.
- [12] Evans, M. and Moshonov, H. (2006). Checking for prior-data conflict. *Bayesian Analysis* **1**, 893–914.
- [13] Friede, T. and Kieser, M. (2001). A comparison of methods for adaptive sample size adjustment. *Statistics in Medicine* **20**, 3861–3873.
- [14] Held, L. and Sauter, R. (2016). Adaptive prior weighting in generalized linear models. *Biometrics* Published online at <http://dx.doi.org/10.1111/biom.12541>.
- [15] Hobbs, B. P., Carlin, B. P., Mandrekar, S. J., and Sargent, D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics* **67**, 1047–1056.
- [16] Hobbs, B. P., Carlin, B. P., and Sargent, D. J. (2013). Adaptive adjustment of the randomization ratio using historical control data. *Clinical Trials* **10**, 430–440.
- [17] Hobbs, B. P., Sargent, D. J., and Carlin, B. P. (2012). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis* **7**, 639.
- [18] Ibrahim, J. G. and Chen, M.-H. (2000). Power prior distributions for regression models. *Statistical Science* **15**, 46–60.

- [19] Ibrahim, J. G., Chen, M.-H., Gwon, Y., and Chen, F. (2015). The power prior: theory and applications. *Statistics in Medicine* **34**, 3724–3749.
- [20] Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2003). On optimality properties of the power prior. *Journal of the American Statistical Association* **98**, 204–213.
- [21] Neelon, B. and O'Malley, A. J. (2010). Bayesian analysis using power priors with application to pediatric quality of care. *Journal of Biometrics & Biostatistics* **1**, 1–9.
- [22] Neuenschwander, B., Branson, M., and Spiegelhalter, D. J. (2009). A note on the power prior. *Statistics in Medicine* **28**, 3562–3566.
- [23] NIST (2016). Digital Library of Mathematical Functions. <http://dlmf.nist.gov/>, Release 1.0.14. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller and B. V. Saunders, eds.
- [24] Pocock, S. J. (1976). The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases* **29**, 175–188.
- [25] Rubinstein, E., Cammarata, S. K., Oliphant, T. H., and Wunderink, R. G. (2001). Linezolid (PNU-100766) versus vancomycin in the treatment of hospitalized patients with nosocomial pneumonia: a randomized, double-blind, multicenter study. *Clinical Infectious Diseases* **32**, 402–412.
- [26] Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D., and Neuenschwander, B. (2014). Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics* **70**, 1023–1032.
- [27] Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley & Sons.
- [28] Viele, K., Berry, S., Neuenschwander, B., Amzal, B., Chen, F., Enas, N., Hobbs, B., Ibrahim, J. G., Kinnersley, N., Lindborg, S., et al. (2014). Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical Statistics* **13**, 41–54.
- [29] Wunderink, R. G., Cammarata, S. K., Oliphant, T. H., and Kollef, M. H. (2003). Continuation of a randomized, double-blind, multicenter study of linezolid versus vancomycin in the treatment of patients with nosocomial pneumonia. *Clinical Therapeutics* **25**, 980–992.

A. Normal Power Prior

As with the binomial setting, it is possible to construct a power prior with a normal likelihood. In the normal setting with an unknown μ , we assume both the historical and current data are realisations of normal random variables and that the parameter of interest is the mean of the current data, so the likelihoods used, $L(\mu; x_0)$ and $L(\mu; x_*)$, are based on normal densities. We set the variances of these distributions, σ_0^2 and σ_*^2 , to the squared standard error of the parameters of interest μ , and so they are treated as fixed in the following derivation. In contrast to the binomial case, many quantities are now available in closed form.

A.1. Power Prior

In the normal setting with unknown mean μ and fixed variance σ_0^2 , the power prior is

$$p(\mu | x_0, \delta) \propto (2\pi\sigma_0^2)^{-\delta/2} \exp\left(-\frac{\delta(\mu - x_0)^2}{2\sigma_0^2}\right) p(\mu).$$

The posterior, based on x_* and σ_*^2 , is

$$p(\mu | x_*, x_0, \delta) \propto (2\pi)^{-(1+\delta)/2} \sigma_0^{-\delta} \sigma_*^{-1} \exp\left(-\frac{(\mu - x_*)^2}{2\sigma_*^2} - \frac{\delta(\mu - x_0)^2}{2\sigma_0^2}\right) p(\mu),$$

so with a flat initial prior for μ we have

$$\mu | x_*, x_0, \delta \sim N\left(\left(\frac{x_*}{\sigma_*^2} + \frac{\delta x_0}{\sigma_0^2}\right) \left(\frac{1}{\sigma_*^2} + \frac{\delta}{\sigma_0^2}\right)^{-1}, \left(\frac{1}{\sigma_*^2} + \frac{\delta}{\sigma_0^2}\right)^{-1}\right).$$

A.2. Full Bayes

The joint power prior based on (3) for μ and δ , assuming a uniform prior on μ and fixed variance, is [21]

$$p(\mu, \delta | x_0) = N(x_0 | \mu, \sigma_0^2/\delta) \text{Be}(\delta | \alpha, \beta).$$

Duan et al. [11] also derive a more general form allowing for unknown variance, which requires bounds on the range of δ depending on the prior on σ_0^2 . The joint posterior distribution is proportional to the joint prior multiplied by the likelihood of the current data, $L(\mu; x_*)$,

$$p(\mu, \delta | x_*, x_0) \propto N(x_* | \mu, \sigma_*^2) N(x_0 | \mu, \sigma_0^2/\delta) \text{Be}(\delta | \alpha, \beta). \quad (12)$$

To obtain the marginal posterior distribution of μ , which is the basis for a fully Bayesian approach, we integrate δ out of (12), so

$$\begin{aligned} p(\mu | x_*, x_0) &\propto \int_0^1 p(\mu, \delta | x_*, x_0) d\delta \\ &\propto \int_0^1 N(x_* | \mu, \sigma_*^2) N(x_0 | \mu, \sigma_0^2/\delta) \text{Be}(\delta | \alpha, \beta) d\delta \\ &= N(x_* | \mu, \sigma_*^2) \frac{1}{B(\alpha, \beta)} \frac{1}{\sqrt{2\pi\sigma_0^2}} \int_0^1 \sqrt{\delta} \text{Exp}\left(\delta \frac{(x_0 - \mu)^2}{-2\sigma_0^2}\right) \delta^{\alpha-1} (1-\delta)^{\beta-1} d\delta \\ &= N(x_* | \mu, \sigma_*^2) \frac{1}{B(\alpha, \beta)} \frac{1}{\sqrt{2\pi\sigma_0^2}} \Gamma(\alpha + 1/2) \Gamma(\beta) M\left(\frac{1}{2} + \alpha, \frac{1}{2} + \alpha + \beta, -\frac{(x_0 - \mu)^2}{2\sigma_0^2}\right) \\ &\propto \exp\left(-\frac{(x_* - \mu)^2}{2\sigma_*^2}\right) M\left(\frac{1}{2} + \alpha, \frac{1}{2} + \alpha + \beta, -\frac{(x_0 - \mu)^2}{2\sigma_0^2}\right), \end{aligned}$$

where $M(a, b, z) = 1/(\Gamma(a)\Gamma(b-a)) \int_0^1 e^{zt} t^{a-1} (1-t)^{b-a-1} dt$ is the confluent hypergeometric function [23, Sec 13.4] which is implemented in standard numerical mathematics libraries. No further estimation of parameters is required and we can use this distribution for inference on μ . Similarly, the marginal posterior of δ is found by integrating μ out of (12), giving

$$p(\delta | x_*, x_0) \propto \begin{cases} N(x_* | x_0, \sigma_0^2/\delta + \sigma_*^2) \text{Be}(\delta | \alpha, \beta) & \text{if } 0 < \delta < 1 \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

This marginal distribution can be used to examine the behaviour of the weight parameter for different values of x_0 and x_* .

A.3. Empirical Bayes

To find (9) in the normal setting, we integrate μ out of the joint likelihood (*i. e.* (12) without the $\text{Be}(\delta | \alpha, \beta)$ prior term) to give the marginal likelihood with respect to δ ,

$$L(\delta; x_*, x_0) \propto N(x_* | x_0, \sigma_0^2/\delta + \sigma_*^2). \quad (14)$$

Since this has the form of a normal likelihood, we can use the invariance property of the maximal likelihood estimate $(x_* - x_0)^2$ of the variance and solve for δ . The value of δ which maximises (14) is

$$\hat{\delta} = \frac{\sigma_0^2}{\max\{(x_* - x_0)^2, \sigma_*^2 + \sigma_0^2\} - \sigma_*^2}, \quad (15)$$

where the \max is required to restrict $\hat{\delta} \leq 1$. Substituting this estimate into the conditional posterior distribution of μ gives the EB posterior distribution,

$$p(\mu | x_*, x_0, \delta = \hat{\delta}) \propto \begin{cases} N(\mu | x_*, \sigma_*^2) \times N(\mu | x_0, (x_* - x_0)^2 - \sigma_*^2) & \text{if } (x_0 - x_*)^2 > \sigma_*^2 + \sigma_0^2 \\ N(\mu | x_*, \sigma_*^2) \times N(\mu | x_0, \sigma_0^2) & \text{otherwise.} \end{cases}$$

The availability of analytical forms for $\hat{\delta}$ and the marginal distributions makes working with EB very fast and easy in the normal setting.

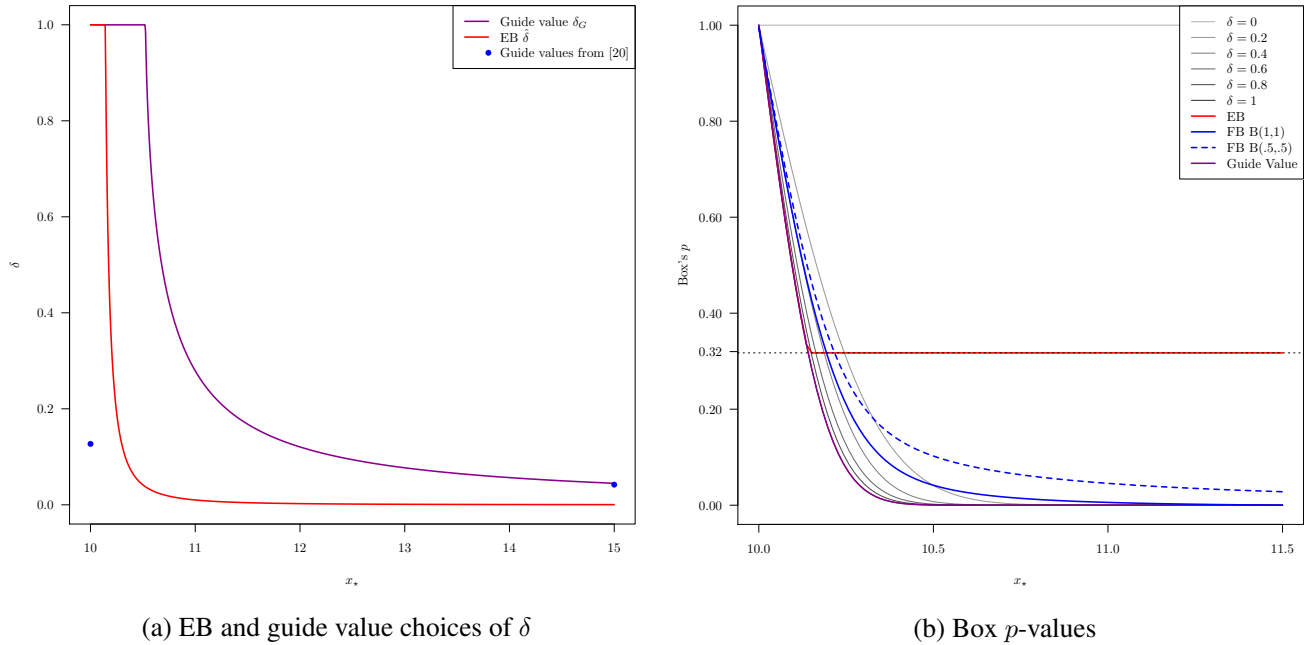


Figure 5. Choices of δ and Box p -values for x_* with $x_0 = 10$ and $\sigma_0^2 = \sigma_*^2 = 1/n_0 = 1/100$. The guide value line in (b) traces exactly the $\delta = 1$ line.

A.4. Guide value

We now examine the guide value in the normal setting and compare the suggested δ_G values with the EB estimate $\hat{\delta}$. Firstly, inserting normal likelihoods into (10) gives

$$h(\delta) = \exp\left(-\frac{(x_0 - x_*)^2}{2(\sigma_0^2/\delta + \sigma_*^2)}\right) \left((2\pi\sigma_0^2)^\delta \sigma_*^2 \left(\frac{\delta}{\sigma_0^2} + \frac{1}{\sigma_*^2}\right) \right)^{-1/2},$$

which is used in $G(\delta)$, which we search numerically for $\delta_G = \arg \min G(\delta)$. Note that due to the use of the unnormalised conditional power prior, $h(\delta)$ includes an extra term involving δ , $(2\pi\sigma_0^2)^{-\delta/2}$, as compared to the marginal likelihood (14).

Ibrahim et al. [20] give an example in a normal setting with varying x_* , fixed $x_0 = 10$, and variances $\sigma_0^2 = \sigma_*^2 = 1/n_0 = 1/n_* = 1/100$, which we replicate here. Figure 5(a) shows the different behaviours of the guide value δ_G and EB $\hat{\delta}$. The greatly increased range of x_* values for which $\delta_G = 1$ as compared to EB is due to the penalty term, $\log(n_0)/\delta$, which penalises smaller values of δ , and therefore increases the minimising value, δ_G . The blue dots show the values of δ_G given in [20], which we were not able to reproduce, although our results do become close for very large differences between x_0 and x_* . For example when $x_* = 50$, both our calculated δ_G and δ_G given in [20] are 0.005.

The values given by Ibrahim et al. [20, table 1 and 2] are shown in Tables 2 and 3. The left side of each table varies the current study size with 100 patients in the historical study and the right side varies the historical size with 100 in the new study.

n	$\delta s_0^2 = 1$	$\delta s_0^2 = 0.5$	n_0	$\delta s_0^2 = 1$	$\delta s_0^2 = 0.5$
5	0.126	0.139	5	0.349	0.378
10	0.127	0.139	10	0.289	0.317
50	0.127	0.140	50	0.166	0.183
100	0.127	0.140	100	0.127	0.140
200	0.128	0.140	200	0.097	0.106
1000	0.128	0.140	1000	0.049	0.054
100000	0.128	0.141	100000	0.006	0.007

Table 2. Guide values for identical historical data and current data, *i. e.* $x_* = x_0 = 10$, current variance $\sigma_*^2 = 1$ and differing historical variance

n	$\delta x_0 = 15$	$\delta x_0 = 50$	n_0	$\delta x_0 = 15$	$\delta x_0 = 50$
5	0.085	0.006	5	0.108	0.014
10	0.060	0.006	10	0.092	0.012
50	0.044	0.005	50	0.054	0.007
100	0.042	0.005	100	0.042	0.005
200	0.041	0.005	200	0.033	0.004
1000	0.041	0.005	1000	0.018	0.002
100000	0.041	0.005	100000	0.006	0.000

Table 3. Guide values for differing historical data and fixed current data and variances, $x_* = 10$, $\sigma_*^2 = \sigma_0^2 = 1$

For all entries in Table 2, the EB values from equation (15) are $\hat{\delta} = 1$, since the data are identical. For Table 3, $\hat{\delta} = 0.04$ when $x_0 = 15$ and $\hat{\delta} = 6 \times 10^{-4}$ when $x_0 = 50$. Under EB the estimates do not depend directly on the size of either the current or historical study, rather only through the variance of the x_0 and x_* . Figure 6 shows the historical data and conditional posteriors using EB and guide values.

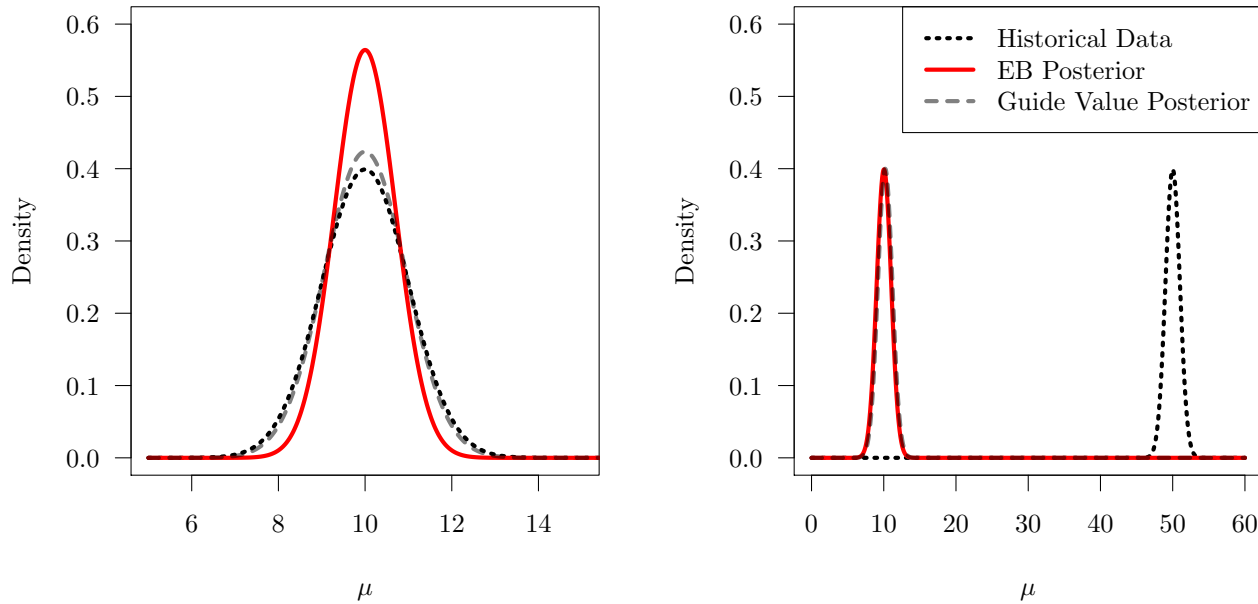
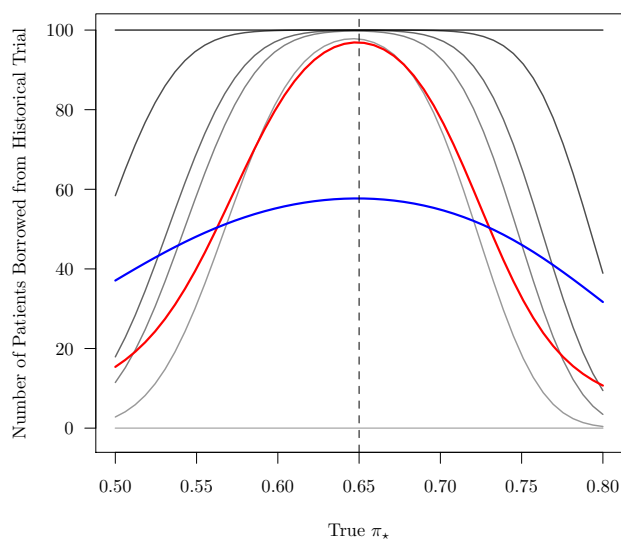


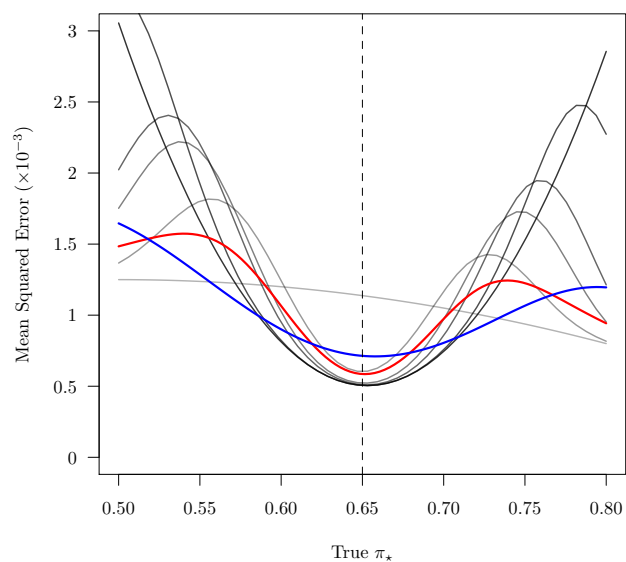
Figure 6. Conditional posterior distribution of μ given δ for $x_0 = 10$ and $x_* = 10$ or 50 with variances $\sigma_0^2 = \sigma_*^2 = 1$ and sample sizes $n = n_0 = 100$.

B. Test-then-Pool Operating Characteristics

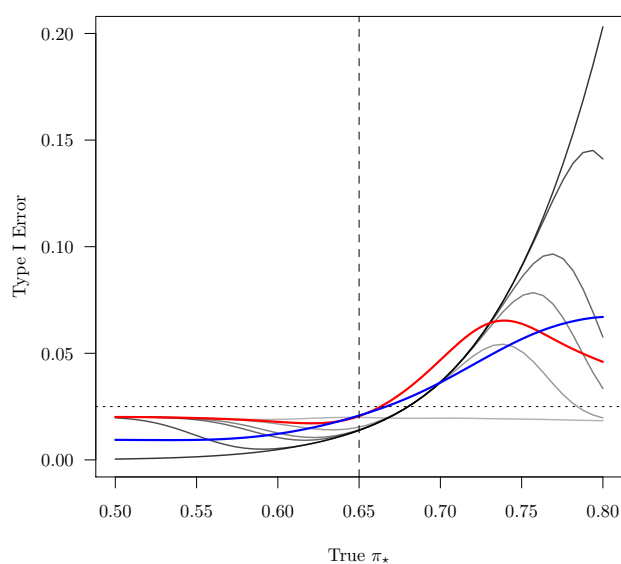
A non-Bayesian alternative that allows pooling dependent on the historical and current data is the test-then-pool methodology [28]. First, the difference between current and historical data is tested at some predefined significance level α , then if the difference is not significant, the data are pooled and if the difference is significant, the historical data is ignored. The significance level used in this testing step need not be the same as that used in the final testing of control versus treatment. When a two-sided Fisher test is used with $\alpha = 0.2$, the all cause mortality data with $p = 0.18$ would ignore x_0 , whereas the clinical cure data which has $p = 0.68$ would pool x_0 and x_* . Such an all-or-nothing approach with a strict but arbitrary cut-off does not seem appropriate and small differences in observed data could lead to very different results. We consider significance levels of $\alpha = 1, 0.2, 0.1, 0.05, 0.01, 0$ for the initial test. The operating characteristics of these test-then-pool and power priors share a similar shape in many cases. The results are shown in Figure 7.



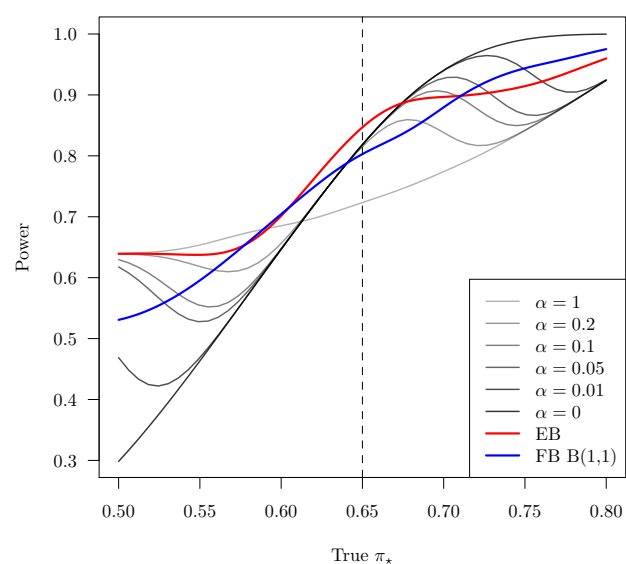
(a) Expected sample size



(b) Mean squared error



(c) Type I error



(d) Power to detect 0.12 difference

Figure 7. Operating characteristics comparing test then pool method with power priors for historical data $\hat{\pi} = 0.65$ and varying true parameter π_* .